DOCUMENT RESUME

ED 257 145 CS 504 903

AUTHOR Roberts, Charles V.
TITLE Preliminary Research Employing the Watson-Barker Listening Test: A Validation of the Instrument.
PUB DATE Mar 85
NOTE 14p.; Paper presented at the Meeting of the International Listening Association (Orlando, FL, March 14-16, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; *Comparative Analysis; *Educational Research; Higher Education; *Listening Skills; Testing Problems; *Test Validity
IDENTIFIERS Receiver Apprehension Test; *Watson Barker Listening Test

ABSTRACT

A study was conducted to further validate the "Watson-Barker Listening Test." The subjects, 120 students enrolled in basic speech courses, completed the Receiver Apprehension Test (RAT) and the Watson-Barker Listening Test: Form A. Statistical analysis of the results revealed a significant correlation between the RAT scores and both the long term memory and the total listening measures on the Watson-Barker test, but not between the RAT scores and the short term memory measure. The results only partially supported claims of validity for the Watson-Barker instrument. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED257145

PRELIMINARY RESEARCH EMPLOYING THE WATSON-BARKER LISTENING TEST: A VALIDATION OF THE INSTRUMENT

Charles V. Roberts

McNeese State University

Lake Charles, Louisiana 70609

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles V. Roberts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A paper prepared for the International Listening Association Convention, March, 1985, Orlando, Florida

CS 504 903

BEST COPY AVAILABLE

ABSTRACT

Preliminary Research Employing the Watson-Barker Listening Test: A Validation of the Instrument

The face validity of the Watson-Barker Listening Test previously has been established through inspection by listening theorists. This study sought additional support for these claims of validity. One hundred twenty students enrolled in basic speech courses were asked to complete the Receiver Apprehension Test (RAT) and take the Watson-Barker Listening Test: Form A. Statistical analysis of the data revealed a significant correlation between RAT scores and both Long Term Memory and Total Listening, but not between RAT scores and Short Term Memory. The significant relationships were curvilinear in nature, as expected, based on the relevant literature. It was concluded that the claims of validity for the Watson-Barker instrument are partially supported by this data.

PRELIMINARY RESEARCH EMPLOYING THE WATSON-BARKER LISTENING TEST: A VALIDATION OF THE INSTRUMENT

The Watson-Barker Listening Test was developed in 1982 in an attempt to create a standardized listening test that would be oriented primarily toward adults and mature college level audiences (Watson and Barker, 1984). While a number of reliability analyses were conducted and acceptable levels of reliability were established, the only measure of validity undertaken was that of "face validity"(Watson and Barker, 1984, p.1). Given the diverse definitions of "listening" held by various listening experts, such support is not totally reassuring. Currently experiments are being conducted in an attempt to link test results of the Watson-Barker instrument with those of other listening tests such as the Kentucky Comprehensive Listening Test. While such experiments will help to establish the efficacy of comparing data of the various tests, they provide only a tautological validation of the instruments. If all tests are highly correlated and if any one test is valid, then the validity claims of all tests can be accepted. If no check of validity other than that of "face validity" is performed, all such claims should be held in abeyance until the concept of "listening" is agreed upon substantively by listening theorists.

The problems of establishing the validity of listening tests are monumental. There is quite a bit of disagreement concerning which various subprocesses should be included within the conceptualization of listening. Is listening a combination of "hearing, understanding, and retaining" information, or should other subprocesses be included or some of these be excluded (Bostrom, 1983)? Regardless of the various conceptualizations, it appears clear from the nature of the instruments

BEST COPY AVAILABLE

being used to measure "listening ability" that the one subprocess that is central to the measurement of listening is the "recall" or "recognition" of retained information. All tests share a common method. Subjects are asked to listen to a message, or set of stimuli, and then are asked to recall or recognize various parts of that message or set of stimuli either immediately after hearing the test passages or at some delayed time thereafter. While the nature of the test passage varies from instrument to instrument, this procedure seems invariant.

Another constant appears to be the effort on the part of the designers to hold "listening motivation" constant for all subjects. All of the major tests of listening ability are administered in such a manner so that all subjects are aware that their listening is to be tested. Kelly (1967) points out the problems of external validity using this procedure when he notes,

> We have a massive body of information about the listening behavior of subjects who knew they were going to be tested. . . but we have done almost nothing to find out about performances across the general range of situations from panic to boredom (p.464).

This is crucial to the external validity of listening tests when one considers that one of the most consistent findings in listening research has been that the recall of material is facilitated by increases in extrinsic motivational cues. Forewarning of a test has been found to be such a cue

Knowledge that a test will follow a listening experience has been labeled "anticipatory set." Anticipatory set creates the real possibility that a "ceiling effect" may be established. Procedures that are common in listening measurement severely limit the free functioning of any antecendent listening ability, as would be manifested in a "non-laboratory" situation. This phenomenon has been reported by many researchers (See, for example, Anastasis, 1961; Kelly, 1962, 1963, 1967,). Cronen and Mihevc (1972) discuss how subjects under "aware" conditions actively listen to messages so that they might answer questions concerning the material at a later time. The effect of forewarning is to

BEST COPY AVAILABLE

raise the motivational forces naturally at work in the typical listener as high as his mental ability will allow and to disallow the differential functioning of other pertinent variables upon the comprehension and retention of material (Kelly, 1967). This may well be the reason that correlations between measures of mental ability and intelligence, and such listening tests as the Brown-Carlsen Listening Comprehension Test and the STEP have been so high (Keller, 1960; Petrie, 1961; Anderson and Baldauf, 1963).

Listening test designers should not be uninterested in studying the listening behavior of subjects under these conditions. Many classroom teachers hope that these conditions exist for them in their various courses. However, even a cursory inspection of the most ideal classroom will reveal that students are not motivated to listen, day in and day out, to the information presented them. Many students seem to be content to remember information only so long as it takes to place that information in their notes. In any case, conditions where testing is immanent are not likely to be found in most other situations.

Of particular interest then is the extent to which scores obtained in controlled conditions of standardized motivation reflect the listening ability of subjects when they venture outside the laboratory environment. Resolving this question of external validity is not an easy task, given the nature of the listening instruments extant today. While the Watson-Barker test does contain stimuli that are capable of being generated in a non-laboratory setting, the task of getting even one subject to respond to questions that would mirror the content of the test under conditions of "nonawareness of the intent to test" is too huge to seriously consider.

Another method is available for establishing the external validity of listening tests. Groups of "good listeners" and groups of "poor listeners" could be given listening tests. In this manner the validity of the instrument could be established. However, before such procedures could be completed, the aforementioned definitional debate as to what

BEST COPY AVAILABLE

constitutes "good listening" would have to be settled.

Bostrom (1984) argues that one method of establishing validity is to illustrate that the instrument in question measures a unique characteristic. He compares a wide variety of tests with his listening instrument to illustrate its uniqueness. While his data is compelling evidence that his instrument measures a unique construct, he presents no evidence that his instrument measures "listening ability." To say that something is not several other things is not the same as saying it is what he says it is. He continues his quest for validity by illustrating that certain groups respond differently than others. Specifically he indicates that college students, army officers, and high school students have different performance levels. Knowing several members of each subject set, I suggest that none of the sets can boast of a uniform level of listening ability. This is not to say that his instrument does not measure listening ability. Rather it is to suggest that he has not substantiated his case for the validity of his instrument in the eyes of this observer.

At least one other method for severing this tautological Gordian Knot was suggested by the efforts of Bostrom (1984). While uniqueness is one characteristic of validity, shared commonality, as evidenced by significant correlations, with valid measures of a phenomenon is acceptable support of a contention of validity. There are tests of established validity that are conceptualized to measure certain aspects of the listening domain. One such instrument is the Receiver Apprehension Test (Wheeless, 1975). This instrument measures the self-reported anxiety of subjects that is associated with listening to stimuli generated in a variety of situations. It has been studied in terms of its relationship to other self-report measures (Beatty, 1981; Beatty and Payne, 1981) and its psychometric properties (Beatty, in press). Of particular note is the established correlation of RAT scores and physiological arousal (Roberts, 1980, 1984). This becomes even more important when the correlation between arousal and retention is entered into the equation. A number of researchers have established a link between retention and arousal

 BEST COPY AVAILABLE

(Kleinsmith and Kaplan, 1963; Crane et al. 1971; Roberts, 1980). The relationship between arousal and retention is posited to be curvilinear in nature, while the realtionship between physiological arousal and RAT scores is linear. Since listening ability is said to reflect short term retention and long term retention ability, in part, then there should be a correlation between RAT scores and scores on valid listening tests. This relationship would be curvilinear in nature. Too much or too little physiological arousal, as indicated by RAT scores, would result in poorer retention scores, as reflected by scores on a listening test. Optimum levels of arousal would result in higher retention scores.

Thus, the following hypothesis was conceived:

> There is a curvilinear relationship between receiver apprehension, as measured by the RAT, and listening ability, as measured by the Watson-Barker Listening Test.

METHOD

SUBJECTS: Subjects were 127 volunteer undergraduate students, 42 males and 85 females, enrolled in beginning speech communication courses at McNeese State University during the Spring semester of 1985. Data of seven of the subjects was subsequently discarded for several reasons. Three of the subjects were from other countries and their grasp of the English language prohibited an accurate test of their listening ability. Four other subjects did not complete one or both of the instruments utilized in this experiment.

PROCEDURE: At the beginning of the Spring semester, students in six sections of a basic speech communications class were asked to volunteer for an experiment. The purpose of the experiment was explained to them in detail and the procedures that would be followed were outlined. They were assured that the tests would have no impact on their grade, nor

BEST COPY AVAILABLE

would their decision to participate or not participate affect their standing in the class. With only one exception, all students agreed to participate. The one non-volunteer was excused from the next class meeting.

At the next class meeting the subjects were asked to complete the Receiver Apprehension Test (Wheeless, 1975). After collecting the RAT, subjects were asked to complete the Watson-Barker Listening Test, Form A (Watson and Barker, 1984). This test requires students to listen to a twenty minute audio tape and answer questions based on the information presented on the tape. There are five different types of listening tasks asked of the subjects. Each section of the test is comprised of ten questions. Three of the sections are said to test "short term memory skills" and the remaining two sections are purported to assess "long term memory skills"(Watson and Barker, 1984). The test tape begins with a short passage that allows the experimenter to insure that all subjects can hear the tape adequately. After adjusting the volume control of the tape player, the tape was played for the subjects, pausing only briefly to allow subjects to turn the pages of their test booklets when required. Although these pauses were not called for in the instructions provided with the test, they were deemed necessary because of the potential for distortion that the extraneous noise presented. The actual test time required varied slightly from class to class (the average time required for completing the Watson-Barker Listening Test was approximately 30 minutes). After the subjects had completed the test, their answer sheets were collected, they were asked to refrain from discussing the tests with others who might subsequently participate in the experiment, and were assured that their test answers would be evaluated , shared and explained to them at the next regular meeting of the class.

RESULTS

The completed tests were scored according to directions provided by the designers of the two instruments. As indicated above, four of the subjects failed to complete one or both of the tests and the tests of three other

BEST COPY AVAILABLE

subjects were discarded because it was evident that they did not understand English well enough to have their listening ability effectively measured by the Watson-Barker instrument. Pearson product-moment correlations were obtained for the scores of the remaining 120 subjects on the RAT and the Watson-Barker test measures of short term memory, long term memory, and total listening ability (short term memory plus long term memory). As suggested by the literature concerning the nature of the relationship between arousal, as tapped by the RAT instrument, and the retention dimension measured by listening tests, no significant relationships were established for total listening ability , short term listening, or long term listening (respectively the results were r=.12, r=.13,r=.06; p>.05).

While a certain level of arousal is necessary to perform cognitive tasks successfully, arousal levels beyond the optimum "readiness" level are dysfunctional (Cofer and Appley, 1964). As indicated above, previous research has shown that there is a significant linear correlation between RAT scores and physiological arousal. A direct relationship between memory and physiological arousal has been established as well. This relationship has been shown to be curvilinear in nature, in line with the "Activation Hypothesis" of Cofer and Appley. Since the Watson-Barker instrument does claim to measure retention, the relationship between it and the RAT most probably would not be linear in nature, but rather would be curvilinear in nature. The further the RAT scores are from the mean RAT score, the lower the Watson-Barker scores should be.

To test this proposed "inverted U-shaped" relationship, the 120 scores were arrayed on a scatter diagram and visually examined. This analysis strongly suggested that the relationship was not linear in nature. To statistically test this relationship the RAT scores of the 120 subjects were converted to absolute scores from the mean of the population (mean=40.89) and Pearson product-moment correlations were obtained for the adjusted RAT scores and the Watson-Barker scores of short term memory, long term memory, and total listening ability (Rosenthal and

BEST COPY AVAILABLE

Rosnow, 1984, pp.222-224). Significant relationships were found to exist between the adjusted RAT scores and long term memory ($r=-.20$, $p<.03$) and between the adjusted RAT scores and total listening ability ($r=-.21$, $p<.02$), but not between the adjusted RAT scores and short term memory ($r=-.12$, $p<.18$). The power of the correlation test was .71 (Cohen, 1977).

DISCUSSION

The hypothesis was supported with regard to the relationships among the RAT scores and both long term memory and total listening ability, but not between short term memory and RAT scores. Previous researchers have suggested a strong link between arousal and long term retention, and a relatively weaker link between arousal and short term retention (Levonian, 1967; Roberts, 1980). These findings are in line with those results. Taken together with the previous literature on the arousal-retention relationship, this study provides evidence for the validity claims of the Watson-Barker Listening Test.

Establishing the validity of any new instrument is difficult. Given the relatively small portion of variance of listening scores that is accounted for by the RAT measure, definitive conclusions concerning the validity of this new instrument must wait for additional data collection. Although the amount of variance accounted for is small, its magnitude is in line with Barker's (1984) conceptualization of listening which posits at least six different subprocesses as being involved with the listening process. "Recall" is only one of these six processes and the only one to which the RAT has been empirically linked. It may well be that recall is of less importance than "attention," "hearing," "understanding," or any of the other possible subprocesses of listening, insofar as total listening scores are concerned.

However, this study does add weight to the claims of external validity for the Watson-Barker instrument. Further testing of the relationship

BEST COPY AVAILABLE

between this listening test and measures of "attention," "understanding," etc., would help to increase confidence in this procedure. A more direct test of the relationship between listening scores on the Watson-Barker test and physiological arousal seems called for as well.

One additional note of caution is called for, based on the research project outlined above. While many claims of "face validity" have been made by the designers of listening tests, most of these tests seem, on the surface, to fail that test of validity because of the single medium nature of the test stimulus. Listeners generally do not "listen" with just their ears. Listening typically takes place while the listener is hearing and viewing the sender of the message. While attempting to assess the listener's ability to analyze the paralanguage message as well as the verbal message is indeed a useful pursuit, neglecting to measure the listener's ability to gain knowledge from the other aspects of nonverbal message transmission may render the total testing procedure useless in terms of applying the results to everyday encounters. Efforts are being undertaken to develop a listening test that more accurately measures the full range of decoding activities that the typical "listening" task involves. This new measurement procedure would include both the aural and the visual stimuli that are present in most communication situations. It is hoped that this new version of the Watson-Barker Listening Test will be found to be an even more valid and reliable measure of that nebulous concept we call listening.

BEST COPY AVAILABLE

BIBLIOGRAPHY

Anastasis, A. Psychological Testing. New York: Academic Press, 1961.

Anderson, H. & R. Baldauf. A study of a measure of listening, Journal of Educational Research, 57 (1963), 197-200.

Barker, L. Communication. Englewood Cliffs, N.J.:Prentice Hall, Inc., 1984.

Beatty, M. Receiver apprehension as a function of cognitive backlog, Western Journal of Speech Communication, 45 (1981), 275-279.

Beatty, M. The effects of anticipated listening anxiety on receiver apprehension scores, Central States Speech Journal, in press.

Beatty, M. & S. Payne. Receiver apprehension and cognitive complexity, Western Journal of Speech Communication, 45 (1981), 363-369.

Bostrom, R. The Kentucky Comprehensive Listening Test. Lexington: Kentucky Listening Research Center, 1983.

Bostrom, R. Research Update. Lexington: Kentucky Listening Research Center, 1984.

Cofer, C. & M. Appley. Motivation: Theory and Research. New York: John Wiley and Sons, 1964.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.

Crane, L., R. Dieker, & C. Brown. Physiological responses to communication, in Listening: Readings, Vol. 2 (S. Duker, ed.), Metuchen, New Jersey: Scarecrow Press, 1971.

Cronen, V. & N. Michevc. Evaluation of deductive argument: A process analysis, Speech Mopnographs, 39 (1972), 124-131.

Keller, P. Major findings in listening in the past ten years, Journal of Communication, 10 (1960), 29-38.

Kelly, C. Actual listening behavior of industrial supervisors, as related to "listening ability,"general mental ability, selected personality factors and supervisory effectiveness, Unpublished Dissertation, Purdue University, 1962.

BEST COPY AVAILABLE

Kelly, C. An investigation of the construct validity of two commercially published listening tests, Speech Monographs, 32 (1965), 139-143.

Kelly, C. Listening: Complex of activities - and a unitary skill? Speech Monographs, 34 (1967), 455-465.

Kelly, C. Mental ability and personality factors in listening, Quarterly Journal of Speech, 49 (1963), 152-156.

Kleinsmith, L. & S. Kaplan. The interaction of arousal and recall interval in nonsense syllable paired-associate learning, Journal of Experimental Psychology, 65 (1963), 190-193.

Levonian, E. Retention of information in relation to arousal during continuously-presented material, American Educational Research Journal, 4 (1967), 103-116.

Petrie, C. An experimental evaluation of two methods for improving listening comprehension abilities, Unpublished dissertation, Purdue University, 1961.

Roberts, C. A physiological validation of the receiver apprehension test, Communication Research Reports, 1 (1984), 126-129.

Roberts, C. The arousal-retention relationship in communication research, Paper presented at the Eastern Communication Association Convention, 1980.

Roberts, C. & T. Steinfatt. Source credibility and physiological arousal: An important variable in the credibility-information retention relationship, Southern Speech Communication Journal, 48 (1983), 340-355.

Rosenthal, R. & R. Rosnow. Essentials of Behavioral Research. New York: McGraw-Hill Book Company, 1984.

Watson, K. & L. Barker. Watson-Barker Listening Test. New Orleans: Spectra, Inc., 1984.

Wheeless, L. An investigation of receiver apprehension and social context context dimensions of communication apprehension, Speech Teacher, 24 (1975), 261-268.

BEST COPY AVAI...